Constructing the effect of alternative intervention strategies on historic epidemics: supplementary appendices

A. R. Cook, G. J. Gibson, T. R. Gottwald, C. A. Gilligan

Appendix A: generating stochastic epidemics via the Poisson and Sellke constructions

The Sellke and Poisson constructions can be used to specify alternative underlying or latent stochastic processes for standard stochastic epidemic models, leading to sampling properties identical to those generated, for example, by Gillespie’s algorithm (Gillespie, 1977). Readers may wish to refer to Svensson (2007) and Sellke (1983) for further discussion of the Poisson and Sellke constructions, respectively. Predicting the effect of mooted interventions operates by making explicit assumptions about how interventions impact on the respective latent processes.

**Poisson construction**

Assume that hosts $i$ and $s$ make contact (sufficient for disease to pass from $i$ to $s$ if $i$ is infected at time $t$ and $s$ susceptible) with each other as a time-inhomogeneous Poisson process (Cox & Isham, 1980) at rate $\beta_{is}(t)$, i.e. that within a vanishingly small time period $(t, t + h)$ the probability of a single contact occurring between them is $\beta_{is}(t)h + O(h^2)$, where $O()$ is the Landau symbol. For homogeneously mixing populations whose contacts patterns are unaffected by time, $\beta_{is}(t) = \beta$ for all $i$, $s$, and $t$. For spatially-extended systems such as the citrus canker example in the main paper, $\beta_{is}(t)$ may be a function of the distance (Euclidean or otherwise) between $i$ and $s$.

Note that although the labels $i$ and $s$ were chosen deliberately to evoke infectious and susceptible hosts, contacts are deemed to occur regardless of
the infectious status of the two hosts. *Infectious contacts*, on the other hand, occur if one host is infected and the other is susceptible.

**Common cold example:** We model contacts between $i$ and $s$ as occurring at rate $\beta_{is}(t) = \beta$ if $t < \tau$ and $\beta \xi$ if $t > \tau$, where $\tau$ is the time of instigation of control.

**Citrus canker example:** We model contacts between $i$ and $s$ as occurring at rate $\beta_{is}(t) = \beta \exp(-d_{is}/\alpha)$, where $d_{is}$ is the Euclidean distance from $i$ to $s$. In addition, infectious contacts from outwith the system occur at rate $\beta_{i\rightarrow s}(t) = \epsilon$.

If the Poisson process is time-homogeneous (i.e. $\beta_{is}(t) = \beta_{is}$), it is trivial to simulate contacts between hosts: these occur with inter-event times being exponentially distributed at rate $\beta_{is}$ (Cox & Isham, 1980). An infinite number of contacts occur between each pair of hosts given an infinite amount of time. Two approaches are to generate contacts only when needed or to generate all contacts up to some specified end-time (perhaps extending the time frame if later required).

If the Poisson process is time-inhomogeneous, contacts may still be generated. If there is a finite number of step-changes (such as in our common cold example), the memoryless property of the exponential distribution can be utilised to generate contacts up until the next step-change, disregard any that follow the step-change and resimulate following the step-change using the new parameter value. If the contact rate varies continuously (see, for example, Cook *et al.*, 2007), then the simulated inter-event time between $t_k$ and $t_{k+1}$ is the solution to $\int_{t_k}^{t_{k+1}} \beta_{is}(t)\,dt = V$ where $V \sim \text{Exp}(1)$ if this solution exists (if it does not exist, then no more contacts occur after $t_k$).

Once the set of contacts has been generated for all pairs of hosts in the population, within host dynamics can be generated independently for each host according to the model. For example, in the SIR model used in our common cold example, post-infection the host becomes infectious immediately and for an exponentially distributed length of time. The transmission of infection can be overlaid on top of this network deterministically, by searching for the next event time and nature, incrementing time, and changing the disease-status of relevant hosts.

Mooted interventions can be effected easily if they change the within host dynamics. For instance, in our citrus canker example, mooted interventions would remove hosts from the infectious category and possibly also from the susceptible category, and truncating the corresponding Poisson processes—
infectious contacts involving the notionally removed hosts—at the time of removal of the respective hosts. Interventions that remove bonds can also be considered. Haydon et al. (2003), for example, consider interventions that remove infectious bonds if the distance between donor and recipient farm is greater than some threshold for some time period. Arbitrary reductions to the contact rate can be effected using Rényi’s splitting theorem (Rényi, 1964; Srivastava, 1971): if the mooted intervention reduces the infection rate from $\beta$ to $\beta\xi$, say (as in our common cold example), then contacts are independently discarded with probability $\xi$. Similarly, interventions that add contacts can also be effected using standard properties of Poisson processes, if so desired.

**Sellke construction**

Associate with each host $s$ a random threshold to infection $Z_s \sim \text{Exp}(1)$. Infectious pressure accrues on $s$ from $i$ at time $t$ at ‘rate’ $\beta_{is}(t)\mathbf{1}\{i \in \mathcal{I}(t)\}$ where $\mathbf{1}\{A\} = 1$ if $A$ is true and 0 otherwise and $\mathcal{I}(t)$ is the set of infectious hosts at time $t$. Infectious pressure is additive over all sources of infection. By time $t$, $s$ has been exposed to cumulative infectious pressure $\int_0^t \sum_i \beta_{is}(u)\mathbf{1}\{i \in \mathcal{I}(u)\} \, du$. The infection of $s$ occurs at the instant this integral equals $Z_s$. If the integral is always less than $Z_s$, then infection never occurs (for example, if all infectious hosts recover first, or temporal forcing stops further infection occurring).

To generate a realised epidemic using Sellke’s construction, begin by simulating a random threshold and the within host dynamics of each host. For each host, calculate a potential infection time that will be the final infection time if no other events occur first. Search for the next event time from the set of potential infection times and within-host event times, increment time, change the disease-status of relevant hosts, and adjust the potential infection time of all relevant hosts. Proceed until no further events occur.

Mooted interventions that change the infection rates are readily incorporated by changing $\beta_{is}(t)$ for some or all $i$, $s$ or $t$; within-host dynamics are similarly readily changed. In our citrus canker example, when trees are removed from the infectious class, the infectious pressure on other hosts decreases and their potential infection times re-evaluated. The interventions considered by Haydon et al. (2003) could be effected by reducing the infection rate to 0 for relevant pairs of hosts at the specified time. In our common cold example, after control is instigated, the infectious pressure on all remaining susceptibles decreases and causes their potential infection times to
be re-evaluated.

Appendix B: fitting the models and constructing notional outbreaks

The use of Markov chain Monte Carlo integration techniques together with data augmentation to fit stochastic epidemic models to partially-observed data is well established; the reader who is unfamiliar with the approach is invited to refer to Gibson & Renshaw (1998) and O’Neill & Roberts (1999). In essence, the main difficulty in fitting epidemic models is that the likelihood (i.e. the probability (mass or density) of observing the data for a given parameter vector) is not tractable except for very small populations (see Ross et al., 2006) for the type of data that are typically observable: for instance partial case notification. If the exact times at which events occur were known, the likelihood could be evaluated. The approach advocated by Gibson & Renshaw (1998) is therefore to treat the unknown event times as unknown parameters, and estimate their distribution jointly with the actual model parameters. This approach has been used for different epidemic models, variously including contact structures, host heterogeneity and within-host dynamics, by inter alios Gibson & Renshaw (1998); O’Neill & Roberts (1999); Gottwald et al. (1999); O’Neill & Becker (2001); Streftaris & Gibson (2004); Höhle et al. (2005); Cook et al. (2007).

In pseudo-code, the approach is:

1. Set $k = 0$. Arbitrarily choose initial values of the parameters $\theta^k_1, \theta^k_2, \ldots, \theta^k_m$ and event times $t^k_1, t^k_2, \ldots, t^k_n$ ensuring that a configuration of both consistent with the data (i.e. with non-zero likelihood), using superscripts to denote iteration.

2. For each parameter, say $\theta_p$, propose a change $\theta^k_p \to \theta^*_p$ through a proposal distribution. Accept this proposal with probability given by the Metropolis-Hastings algorithm, setting $\theta^{k+1}_p = \theta^*_p$; if rejected, set $\theta^{k+1}_p = \theta^k_p$. Increment $k$.

3. For each event time, say $t_p$, propose a change $t^k_p \to t^*_p$ through a proposal distribution. Accept this proposal with probability given by the Metropolis-Hastings algorithm, setting $t^{k+1}_p = t^*_p$; if rejected, set $t^{k+1}_p = t^k_p$. Note that non-valid configurations (e.g. a host recovers from
infection before it is infected) may be automatically rejected as the likelihood is zero. Increment $k$.

4. Repeat the last two steps until the Markov chain is deemed to have converged, discarding an initial burn-in period.

See Gilks *et al.* (1996) and references therein for an excellent introduction to MCMC sampling and for details of the Metropolis-Hastings algorithm.

At any iteration $k$ in the MCMC routine, we have a draw $\theta^k_1, \theta^k_2, \ldots, \theta^k_m$ and $t^k_1, t^k_2, \ldots, t^k_n$ from the joint posterior distribution of parameters and event times. These can easily be converted to $\hat{\theta}^k_1, \hat{\theta}^k_2, \ldots, \hat{\theta}^k_m$ and Sellke thresholds $\hat{Z}^k_1, \hat{Z}^k_2, \ldots, \hat{Z}^k_n$ by reversing the reconstruction routine described in Appendix A. The effect of a mooted intervention can therefore be reconstructed by converting from actual (inferred) times to Sellke thresholds and then back to notional times that incorporate the effects of the mooted intervention. Doing this for multiple iterations from the MCMC routine allows a distribution to be built-up that fully takes account of the parametric uncertainty (through the posterior distribution for the parameters) as well as trajectory uncertainty (representing the uncertainty in the infection times).

The description above assumes that all hosts are infected. Should one or more hosts not be infected (and have no corresponding infection time), the associated Sellke threshold is not specified precisely. Rather, there is a distribution of Sellke thresholds consistent with the non-infection of the host. This distribution is an $\text{Exp}(1)$ distribution conditioned on the minimum value of the threshold such that no infection would occur (were the threshold smaller than this minimum, then the host would have been infected after all). The easiest way to account for this in data with a final observation time $T$ is to calculate the integrated infective challenge on the uninfected host $s$ until $T$ and add a randomly generated $\text{Exp}(1)$ variate. Note that this is fully consistent with the model, the data and the likelihood principle.

**Data and Code**

The common cold data and C++ code for jointly fitting the model and evaluating the effect of mooted interventions on this system are available on-line in compressed Tar file format, from which the original files may be recovered in Unix or Linux using the command `tar zxvf cggg08_common_cold.tar.gz` in the directory containing the file.
References


